速变服务网络行为特征分析①

褚燕琴[1,2], 应凌云[1], 冯登国[1], 苏璞睿[1]

[1](中国科学院软件研究所, 北京 100190)
[2](中国科学院大学, 北京 100049)

摘 要: 速变(Fast-Flux)服务网络通过返回不断变化的DNS解析结果, 以大量被攻陷主机的IP地址作为服务地址, 利用被攻陷主机进行重定向形成服务网络. 长时间跟踪并记录了分布在全球6大洲的300个DNS服务器对23000多个域名的解析结果. 根据DNS解析数据, 文章对比其他研究成果从多个维度对Fast-Flux恶意域名和Fast-Flux服务网络的行为特征进行了全面讨论, 并进一步开展特征辨识度分析. 关于Fast-Flux服务网络行为特征的分析揭示了FFSN新的行为变化, 为恶意域名检测、网络资源保护等提供了依据.

关键词: Fast-Flux 服务网络; Fast-Flux 域名; 行为特征; 行为分析; DNS

Behavioral Analysis of Fast Flux Service Network

CHU Yan-Qin[1,2], YING Ling-Yun[1], FENG Deng-Guo[1], SU Pu-Rui[1]

[1](Institute of Software, Chinese Academy of Sciences, Beijing 100190, China)
[2](University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Fast-Flux Service Network is a network of compromised computer systems with public DNS records that are constantly changing. In this paper, we collected DNS data from 300 different domain servers spanning 6 continents. Our engines worked for over 1600 hours on more than 23000 domains. With the tremendous DNS reply data, we are able to analyze the behavior of Fast-Flux Domain and Fast-Flux Service Network through different dimensions, and locate the most distinguishing features. This paper helps to reveal new features compared with previous researches, and verify the dynamic nature of Fast-Flux Service Network. Behavioral analysis of Fast-Flux Service Network also provides a view for malicious domain detection and network resource protection.

Key words: Fast-Flux Service Network; Fast-Flux domain; behavioral feature; behavioral analysis; DNS

网络服务的质量很大程度上取决于服务的可用性和响应速度. 为了确保服务的高可用性和高响应度, 服务提供商在世界各地建立数据中心, 一方面, 分布式处理客户请求, 另一方面, 建立冗余服务网络用于处理服务故障. 服务提供商利用DNS解析, 使服务请求更合理地导向响应数据中心. CDNs(Content Distribution Networks)和RRDNS(Round-Robin DNS)为负载均衡和服务冗余提供了方法.

恶意服务控制者如钓鱼网站、恶意代码传播网站、非法药品销售网站等同样需要设法保持其服务长期可用. 但是, 此类服务网站面临一个问题, 域名或IP一旦被确认提供恶意服务即可能立刻被限制. 因此, 该类网站需要新的技术确保服务可用性.

Fast-Flux 服务网络(Fast-Flux Service Network, FFSN)是近几年出现, 在恶意服务中得到广泛应用的一种网络类型, 其采用Fast-Flux技术逃避IP侦测保证服务在线. 随着网络的发展, FFSN 随之发展变化, 分析 FFSN 新的行为特征为恶意域名检测、网络资源保护等提供了依据.

1 FFSN概述

FFSN 不断变化 DNS 解析结果, 以大量被攻陷主

① 基金项目:国家自然科学基金(61073179);北京市自然科学基金(4122086);国家科技重大专项项目(2011ZX03002-005-02);国家重点基础研究计划(973)(2012CB315804)
收稿时间:2013-01-31;收到修改稿时间:2013-02-28

© 中国科学院软件研究所 http://www.jos.org.cn

机的 IP 作为域名服务 IP[1], 从而避免 IP 检测限制等导致的服务不可用. FFSN 和 CDNs 及 RRDNS 存在相似之处. CDNs 和 RRDNS 利用 DNS 服务向客户返回多个 IP 地址, 返回 IP 就近选择、多次轮转, 从而避免单个失效 IP 导致的服务不可用[2,3]. FFSN 一次返回多个随机 IP, 多次请求返回的 IP 地址不断变化. FFSN 返回 IP 的随机性和变化性主要在于: 1)FFSN 控制大量主机, 对应一个庞大的 IP 地址池, 当客户请求 DNS 解析时, FFSN 通过一定的策略返回 IP 地址池中若干 IP 的集合; 2)IP 地址池根据被控制主机变化. 利用 Fast-Flux 托管技术, 恶意服务控制者将控制网络变成一个全球范围的服务网络, 被控制主机作为 flux 代理重定向客户请求. FFSN 攻陷并控制大量主机构建服务, 对网络资源造成了巨大的消耗; 托管于 FFSN 的 Fast-Flux 恶意域名(Fast-Flux Domain, FFD)提供的服务通常是违法的, 其内容的传播对社会秩序造成了影响. 因而, 研究 FFSN 的行为特征具有现实意义.

针对 FFSN 的研究最早出现于 2006 年. Honeynet 项目组 2007 年发表报告对 FFSN 进行了系统的介绍[1]. Holz 等[4]对 FFSN 开展了试验性研究. 研究对比分析了托管于 FFSN 和 CDNs 的域名行为, 对 FFSN 行为特征进行了讨论和总结. Passerini 等[5]分析总结了 FFSN 的 9 个特征, 包括域名注册时间、域名注册商、域名 A 记录、域名 TTL 值等, 并将其应用于 FFD 检测. 但是, 其所选取的部分特征如 A 记录数、TTL 值等在当前的 FFSN 应用中发生了变化. Si-Yu 等[6]在文中提出了新的特征分析角度, 讨论了完全基于地理分布的 FFD 检测方法. 该方法引入两个地理变量作为 FFSN 的关键特征——IP 地址地理分布和服务地理距离. 该方法假设 FFSN 的 flux 代理地理分布松散, 其服务请求者和提供者间距离大于正常域名. 但是在我们的分析中发现单独依赖地理因素标识网络并不可靠. Xin Hu 等[7]结合传统特征和地理分布特征针对 FFSN 在全球范围内 IP 地址的使用进行了分析, 提出了基于支持向量机(Support Vector Machine, SVM)的 FFD 分类器. 另外, Caglayan 等[8]对 FFSN 的生命周期、大小、增长趋势进行了观察.

国内 FFSN 的研究最早出现于 2009 年, 汪洋[9]在硕士论文中开展了 FFSN 相关研究, 论文选取 A 记录数、IP 分散度、TTL 值、域名创建时间四个特征构成检测向量, 采用神经网络和 SVM 用于 FFD 检测. 国内 FFSN 的相关研究相对国外出现较晚, 研究相对较少, 研究主要着重于 FFD 检测技术, 针对 FFSN 的基础研究如特征分析、新特征的挖掘等方面相对较少.

目前, 针对 FFSN 的相关研究主要是面向 FFD 检测的行为特征分析, 不能全面地呈现 FFSN 的行为特征以应用于网络资源保护等其他方面. 同时, 随着 FFSN 的普及以及网络的发展, FFSN 也在发生明显的变化. 本文从多个角度对 FFSN 的特征进行全面的分析, 深入探讨 FFSN 的特征变化.

本文的主要贡献总结如下:

1) 向分布在全球的 300 个域名服务器发起域名解析请求, 对 23000 多个域名进行长时间的跟踪记录, 以确保足够的数据量开展针对 FFSN 的行为特征研究;

2) 从多个维度对 FFSN 行为特征进行分析和总结, 对比以往研究详细探讨 FFSN 特征变化;

3) 构建 FFSN 特征向量, 对分析获取的行为特征进行辨识度分析, 进一步揭示 FFSN 新的变化.

2 实验数据采集

FFSN 将服务映射到分布于全球的 IP 地址, IP 分布相较于正常服务存在较大的差别, 因而面向单一的 DNS 本地服务器请求解析数据不能全面反映 FFSN 的行为特征. 实验选取分布在全球的 300 个 DNS 服务器作为数据采集点, 如表 1, 定时发起 DNS 解析请求.

表 1 DNS 查询节点洲域分布

洲	亚洲	欧洲	北美洲	南美洲	非洲	大洋洲	其他
DNS 数	146	63	62	17	5	5	2

FFSN 被广泛应用于钓鱼网站等, 且垃圾邮件是 FFD 传播的主要方式之一. Holz 指出垃圾邮件中包含的域名 30%以上利用 FFSN 进行托管[4]. 实验从表 2 所示网站搜集了 23000 多恶意域名进行跟踪记录, 并定期从三个网站搜集新的恶意域名加入请求队列. 为了后期比对分析, 我们选取 alexa.com 网站统计认定的全球流量前 100 域名作为对比观测数据.

表 2 域名来源说明

类别	网址	备注
垃圾邮件	http://untroubled.org/spam/	前期搜集
钓鱼网站	http://www.phishtank.com	前期搜集
恶意域名	http://www.abuse.ch/	前期搜集
恶意域名	http://www.malwaredomains.com http://atlas.arbor.net/summary/fastflux https://zeustracker.abuse.ch	定期搜集
正常域名	alexa.com	前期搜集

© 中国科学院软件研究所 http://www.jos.org.cn

实验开发基于 dig 的采集器定时向 300 个域名服务器对 23000 多个域名发起分时并发解析请求, 记录解析结果. 记录的解析结果包括 A 记录、NS 记录、NA 记录等. 由于域名数较多, 域名不断新增, 实验定时统计域名关联 IP 总量(如未特别说明, 下文中关联 IP 均指解析结果 A 记录中对应 IP), 合理分布请求. 对 IP 总量排名前 1000 的域名每分钟发起并发解析请求; 前 5000 的域名每五分钟发起并发解析请求; 其余 20 分钟发起并发解析请求; 新增域名快速迭代并分类. 具体请求调度算法如代码清单 1.

代码清单 1

```
start() {
  #处理 IP 总量前 1000 的域名
  ./dnsQueryDigParalell.sh data/top1k data/digdata 1 &> top1k.log &
  #处理 IP 总量前 5000 的域名
  ./dnsQueryDigParalell.sh data/top5k data/digdata 5 &> top5k.log &
  #处理新增域名
  ./dnsQueryDigParalell.sh data/fresh data/digdata 5 &> fresh.log &
  #处理其他域名
  ./dnsQueryDigParalell.sh data/rest data/digdata 20 &> rest.log &
}
```

以上四种类别内部由子进程并发独立发起 DNS 解析请求, 请求调度算法如代码清单 2.

代码清单 2

```
#并发处理逻辑块
for i in `cat $1`                    #$1 存储待请求解析域名列表
  do
    #选定 DNS 服务器, 服务器轮训请求
    local sdns=${dns[$idx]}
    #发起解析请求且并发处理下一条
    cx $sdns $i >> $2/$i.data &      #$2 存储解析结果文件夹
    #idx 的解耦选择, $mdix 存储数组长度
    idx=$(( ( $idx + 1 ) % $midx ))
    #发起子进程间隔时间, 避免同时请求对 DNS 服务器的压力
    # local slpT=`echo "scale=4; $3*60/$fline - 0.0047" | bc
    sleep $slpT
done
#请求解析函数, 在并发处理逻辑块中调用
cx(){ #在$1 所示 dns 服务器上查询$2 所示 domain
  dig +nocmd +nocomments +nostats +noquestion @$1 $2
}
```

实验记录从 2012 年 3 月 6 日至 2012 年 5 月 11 日连续 1600 小时的 DNS 解析结果. 为了后期分析需要, 每六小时统计域名关联 IP, 对关联 IP 进行可达性检测.

由于观测条件限制和客观网络原因, 实验采集数据可能存在以下噪点:

1) 本文定期检测 DNS 服务器可达性, 将故障服务器替换为同大洲其他有效 DNS 服务器. 实际失效时间与检测到的失效时间存在误差;

2) IP 可达性检测从中国大陆服务器发起, 由于客观网络原因, 存在一定误报.

3 行为特征分析

FFSN 作为恶意网络具有其固有的行为模式, FFD 在一定程度上表征 FFSN 的行为特征. 我们对比历史研究根据实验采集数据从多个维度对 FFD 及 FFSN 行为特征展开全面分析, 对 FFSN 新的行为变化进行详细讨论.

3.1 短期行为分析

1) 单次解析结果 A 记录数

FFSN 的初期研究结果表明 FFD 单次域名解析请求的返回结果中 A 记录个数(5 个以上)一般大于 CDN 域名(3 个左右)或其他域名[4]. 随着网络发展, 正常网络服务提供商扩大服务建设, 单次 DNS 解析请求返回 A 记录数增加; 而有些 FFSN 为了躲避检测减少单次返回 A 记录数. 图 1 显示 google.com 一次域名解析结果中包含 11 条 A 记录, Fast-Flux 域名 datetop.ru 仅返回 2 条 A 记录. 由此可见, 单次解析结果 A 记录的数量已经不能成为区分 FFD 和正常域名的主要特征.

2) 单次解析结果关联 IP 分布

自治系统(Autonomous System, AS)是共享公共路由选择策略管理下的网络集合, 不同网段可能在地理位置上靠近从而属于同一个自治域, 使用相同的路由

© 中国科学院软件研究所 http://www.jos.org.cn

策略[10]. 观察单次返回结果中 A 记录的关联 IP 可知, 正常域名 IP 段比较集中, 而 FFD 比较分散. 为了进一步分析关联 IP 分布特征, 我们利用 whois 查询关联 IP 地址所属自治域及所属国家. 图 1(a)google.com 解析结果 11 条 A 记录关联 IP 均来自同一自治域, 所属国家均为美国; 图 1(b)datetop.ru 的 2 条记录 IP 分别来自不同自治域和国家, 如表 3 所示.

;; ANSWER SECTION:

google.com.	24	IN	A	74.125.227.72
google.com.	24	IN	A	74.125.227.67
google.com.	24	IN	A	74.125.227.65
google.com.	24	IN	A	74.125.227.73
google.com.	24	IN	A	74.125.227.70
google.com.	24	IN	A	74.125.227.78
google.com.	24	IN	A	74.125.227.69
google.com.	24	IN	A	74.125.227.71
google.com.	24	IN	A	74.125.227.68
google.com.	24	IN	A	74.125.227.66
google.com.	24	IN	A	74.125.227.64

(a) google.com

;; ANSWER SECTION:

datetop.ru.	300	IN	A	109.160.104.133
datetop.ru.	300	IN	A	89.237.12.230

(b) datetop.ru

图 1　单次 DNS 解析请求返回 A 记录

表 3　域名 A 记录 IP 地址分布

A 记录中 IP 地址	AS 号	所属国家
google.com		
74.125.227.72	15169	US
74.125.227.67	15169	US
74.125.227.65	15169	US
74.125.227.73	15169	US
74.125.227.70	15169	US
74.125.227.78	15169	US
74.125.227.69	15169	US
74.125.227.71	15169	US
74.125.227.68	15169	US
74.125.227.66	15169	US
74.125.227.64	15169	US
datetop.ru		
109.160.104.133	39251	BG
89.237.12.230	28745	RU

我们对域名进行查询验证发现, FFD 单次解析结果 A 记录中关联 IP 地址所属 ASN 和国家较正常域名分散度高.

3) 单次解析结果 IP 可达性

正常服务提供商确定服务器失效后及时切换有效服务器 IP 提供服务. 正常网络服务提供商对 DNS 返回 IP 地址的有效性具有可控性. 相反, FFSN 为其提供重定向的 flux 代理由被攻陷主机组成, 不能保证 flux 代理始终在线. FFD 单次解析结果返回的 A 记录中关联 IP 的可达性和正常域名存在明显差异. Knysz 在研究中指出 Fast-Flux 服务在和正常服务的博弈中不断改进 IP 分发算法以尽可能保证返回代理 IP 的有效性[11]. 但是, 由于 FFSN 控制代理存在客观缺陷, 实际观测中我们发现 FFD 关联 IP 不可达的情况仍然比较普遍. 我们对多个域名单次解析结果进行 IP 可达性测试发现 FFD 对应 IP 可达性远远小于正常域名.

实际实验中, 一方面由于单次解析结果 IP 可达性检测存在一定误报率; 另一方面对单次解析结果进行实时可达性测试对系统压力较大, 我们采用定时检测方法对域名关联 IP 进行可达性测试. 域名关联 IP 有效性在 3.2 中进行详细讨论.

4) 多次解析结果新 A 记录数

FFSN 通过不断攻陷新的主机作为 flux 代理提供服务. FFSN 设法将新被攻陷主机的 IP 尽快传播出去为其提供服务. 我们对域名连续解析结果进行观察, 正常域名单位时间内出现新域名的几率小于 FFD. 以 google.com 和 datetop.ru 为例, 连续多次解析结果中 google.com 仅对图 1(a)中 A 记录进行轮转, datetop.ru 出现新关联 IP, 且 A 记录数量和 IP 分布均发生改变, 如图 2 所示.

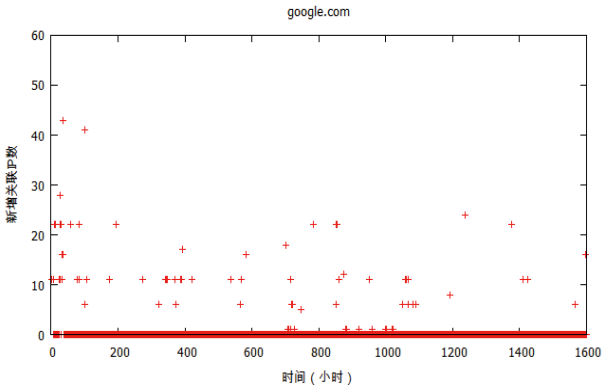
;; ANSWER SECTION:

datetop.ru.	300	IN	A	84.234.6.108	new
datetop.ru.	300	IN	A	89.237.12.230	
datetop.ru.	300	IN	A	93.105.17.49	new

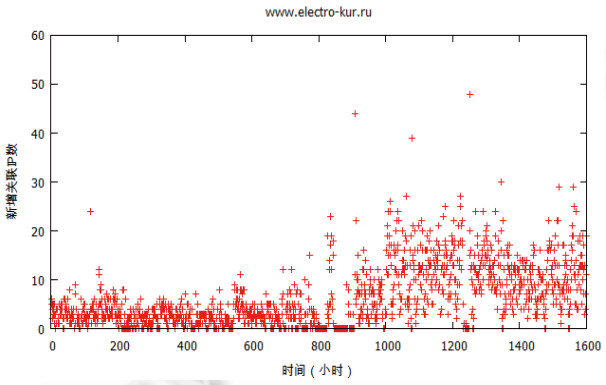
图 2　datetop.ru 第二次 DNS 解析请求返回的 A 记录

本文进一步对域名单位时间内出现新关联 IP 数目进行统计发现, 域名存活期内 FFD 出现新关联 IP 的几率远远大于正常域名. 如图 3 所示, FFD 恶意域名 www.electro-kur.ru 在 1600 小时观测期中出现新关联 IP 的单位时间占总观测时间的 1392/1600; 而 google.com 仅占 82/1600.

© 中国科学院软件研究所 http://www.jos.org.cn



(a) 正常域名



(b) Fast-Flux 域名

图 3 单位时间内新关联 IP 数统计

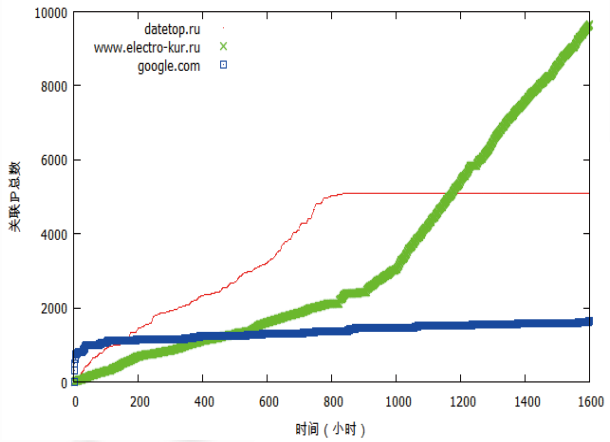
3.2 长期行为分析

1) 关联 IP 总数和增长率

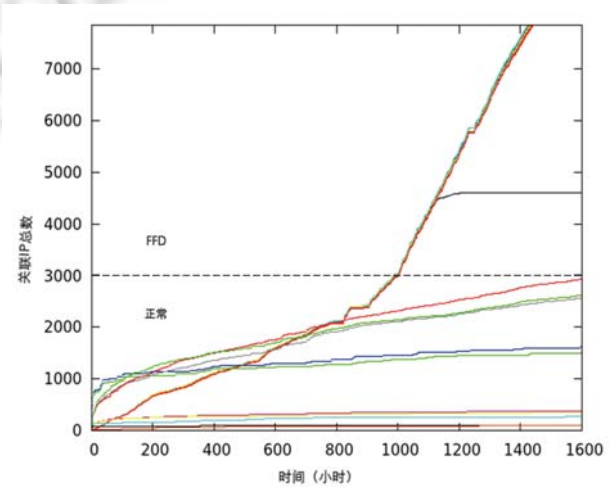
FFSN 不断攻陷新的主机将其 IP 加入到服务 IP 地址池，在接收 DNS 解析请求时从 IP 地址池选取 IP 集合返回给请求客户端. 一个利用 FFSN 提供服务的 FFD 在一段时间内与其关联的 IP 数不断增长.

FFD 关联 IP 总数的相关特性在 Holz、Passerini 等人[4,5]的研究工作中均有讨论. 假设一个 FFD 永远未被检测并限制，其关联 IP 总数总能超过正常域名. 但是实际网络中，一方面 FFD 可能被检测并限制，另一方面攻击者为了服务持续有效会定期更换新的域名提供服务，因此 FFD 存在生命周期，一段时间后域名不再有效；另一方面，正常服务网络不断扩大建设，提供服务的 IP 不断增加. 在我们的观测数据中，最突出的 FFD 存在 2 万多个关联 IP 地址，关联 IP 最少的 FFD 拥有 442 个 IP；比较突出的正常域名拥有将近 7000 个关联 IP. 由此可见，关联 IP 总量已经不能成为区分 FFD 和正常域名的关键特征.

根据 3.1 节讨论的域名单位时间新关联 IP 数统计及域名关联 IP 增长率计算，我们发现 FFSN 关联 IP 增长率相较正常域名存在明显差异. 图 4(a)展示了三个不同域名所关联 IP 的增长状况，其中 google.com 关联 IP 总量在前期快速增长后逐渐趋于平缓；www.electro-kur.ru 整体维持线性增长；datetop.ru 于观测期第 824 小时下线，下线前关联 IP 总数呈线性增长. 图 4(b)中，我们分别对十个 Fast-Flux 域名和十个正常域名关联 IP 增长情况进行统计，图中 1600 小时后总量低于 3000 为正常域名，前期快速增长后增长幅度均维持在较低水平且各域名关联 IP 总量各不同；总量在 3000 以上的为 FFD，关联 IP 数总体呈线性增长，不同域名间增长量及 IP 总量出现了较大的重叠. 该现象为我们后期 FFSN 集群效应分析提供了依据.



(a) 特定域名对比



(b) FFD 与正常域名对比

图 4 域名关联 IP 总量统计

© 中国科学院软件研究所 http://www.jos.org.cn

2) 关联 IP ASN 和国家分布

正常域名的服务器一般在地理上相对集中，IP 地址来自一个或若干个地区; flux 代理是被攻陷主机，分布地点没有特定规律性.

3.1 短期特征分析中，我们得出一次 DNS 解析结果 FFD 返回记录关联 IP 所属自治域较正常域名分散. 我们统计 1600 小时的观测数据对域名关联 IP 所属自治域和国家进行分析.

本文选取总 IP 数相近的两个域名进行比较，其中 ricalt.com 拥有关联 IP 数 2278 个，i.dell.com 拥有关联 IP 数 2592 个. 我们通过 whois 查询获知 ricalt.com 的 2278 个 IP 地址分别来自 596 个不同自治域, i.dell.com 的 2592 个 IP 地址分别来自 177 个自治域. 进一步观察分析, ricalt.com 中仅一个 IP 对应一个自治域的记录多达 314 个; i.dell.com 关联 IP 对应的 177 个自治域中，一个 IP 与之对应的仅有 9 个，如图 5 所示. 一个自治域仅对应一个 IP 的情况，FFD 远远大于正常域名. 我们对 IP 地址所属国家进行统计发现，国家分布拥有和 ASN 分布同样的特征.

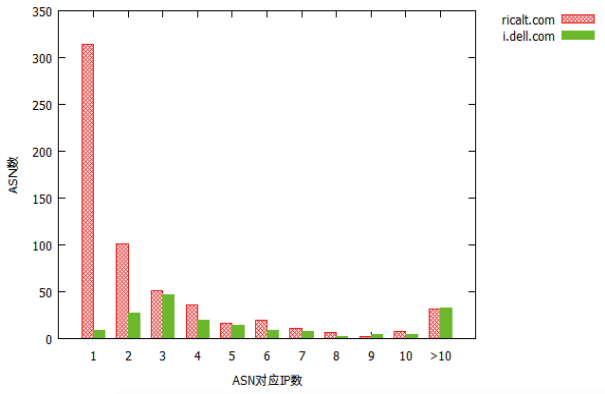
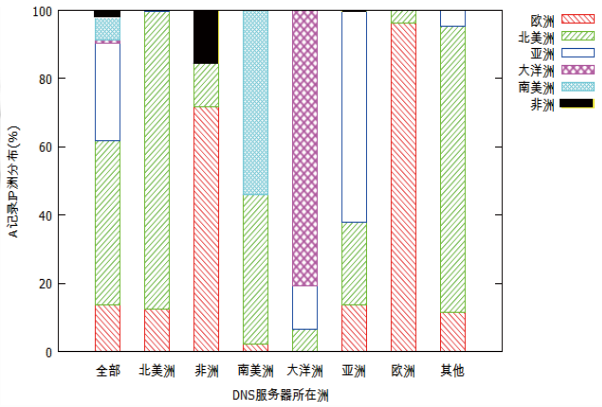


图 5　域名关联 IP 自治域分布统计

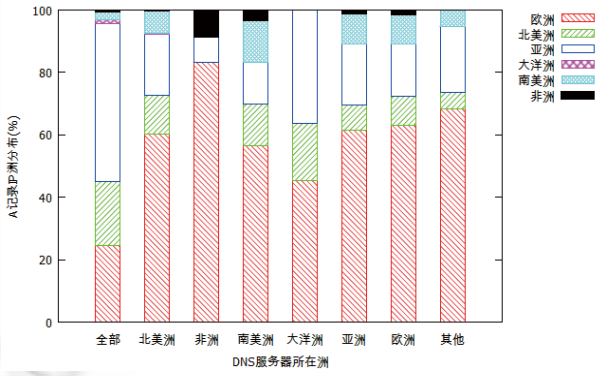
3) 关联 IP 洲分布

正常服务通过 CDN 对客户发起域名解析时采用就近原则，当服务提供商在多个地区设有服务器时，就近返回发起请求的 DNS 服务器所在地的服务器 IP. 正常域名解析结果随发起地变化明显. 本文数据采集自不同洲的 DNS 服务器，为了更准确地分析 IP 地址的地域分布，我们根据 DNS 服务器所在洲对域名关联 IP 进行分类统计，即根据 DNS 服务器所属洲统计各个域名 DNS 解析结果中 A 记录对应 IP 地址的地域分布.

图 6 对 i.dell.com 和 ricalt.com 关联 IP 进行了地域分布统计，第一列标识域名所有关联 IP 的地域分布，其他各列以 DNS 服务器所在洲划分解析结果并统计解析结果中 IP 地址的洲分布. 图 6(a)柱状图显示，正常域名关联 IP 地址的洲际分布随发起请求 DNS 服务器所在地不同出现较明显的变化，如分布在大洋洲、南美洲、非洲的 IP 地址仅在发起地出现且所占比例明显. 而图 6(b)中显示 FFD 关联的 IP 分布相对稳定，并未受地理因素明显影响.



(a)　i.dell.com



(b)　ricalt.com

图 6　域名关联 IP 洲分布统计

4) 关联 IP 有效性

在 3.1 节单次解析结果 IP 可达性中，我们分析了 FFD 单次解析结果 IP 有效性特征. 分析表明，FFD 关联 IP 有效性不可控, IP 失效率较大. 本小节，我们统计分析整个观测期域名关联 IP 的有效性特性.

我们在观测期间对域名所有关联 IP 每六个小时进行一次可达性测试. 测试发现，正常域名关联 IP 的有效数趋近于总关联 IP 数，而 FFD 的有效关联 IP 数远远小于总关联 IP 数. 随着时间的推移, FFD 有效 IP 增长率小于总关联 IP 增长率.

© 中国科学院软件研究所 http://www.jos.org.cn

图 7 中, google.com 的有效关联 IP 数总体趋近于当前关联 IP 数. 由于客观网络原因, 图中出现若干噪点. datetop.ru 的有效关联 IP 数不到总关联 IP 数的 1/3, 并且随着时间的推移总 IP 数和有效 IP 数距离不断拉大. 观测期最后数据显示 datetop.ru 有效关联 IP 数仅为总关联 IP 数的 1/5.

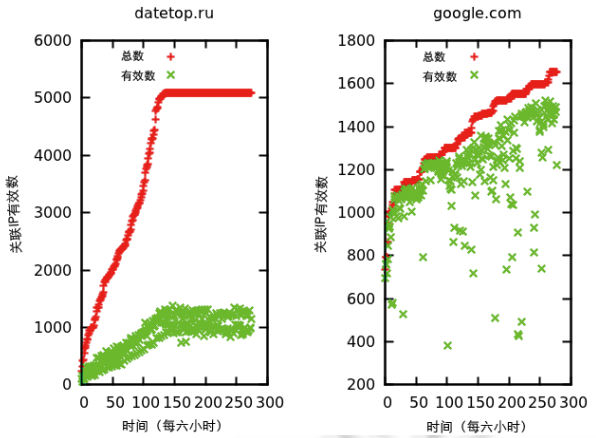


图 7 域名关联 IP 有效数统计

5) 关联 IP 重叠度

FFSN 为了充分利用被攻陷主机的 IP 地址常常会重用 IP. 我们对 FFD 关联 IP 进行重用性分析. IP 重用主要体现在两个方面: 1)域名 A 记录对应 IP 和 NA 记录对应 IP 的重叠度; 2)不同域名 A 记录 IP 地址的重叠度.

网络中正常服务提供商可能采用多个域名作为服务入口, 图 8 www.baidu.com 和 www.content4ads.com 域名解析记录完全重叠. 因而我们在分析中未对不同域名 IP 地址重叠度进行统计和分析.

```
;; ANSWER SECTION:
www.baidu.com.      5    IN    CNAME    www.a.shifen.com.
www.a.shifen.com.   5    IN    A        119.75.217.56
www.a.shifen.com.   5    IN    A        119.75.218.77

;; ANSWER SECTION:
www.content4ads.com.    5    IN    CNAME
zsmain.a.shifen.com.
zsmain.a.shifen.com.    5    IN    A    119.75.217.56
zsmain.a.shifen.com.    5    IN    A    119.75.218.77
```

图 8 正常域名 IP 重叠度示例

对域名的 A 记录和 NA 记录 IP 重叠度进行统计和分析发现, 正常域名 IP 重叠度趋于 0 而 FFD 的 IP 重用度非常高, 如表 4 所示. 其中域名对应 A 记录关联 IP 数为#A, 域名对应 NA 记录关联 IP 数为#NA, A 记录与 NA 记录 IP 重叠数为#overlap.

表 4 域名关联 IP 重叠度统计示例

域名		#A	#NA	#overlap
FFD	dating-trah.ru	9237	10636	5752
	smoke-fresh.ru	10199	3516	2301
正常域名	google.com	1652	28	0
	www.visa.com	2637	7031	1

3.3 其他行为研究

1) 域名注册信息

在前面的分析中我们指出恶意域名存活时间较短. 恶意域名一旦被检测就可能被限制, 恶意域名控制者必须重新注册新的域名提供服务; 或者为了服务连续性在旧的域名被限制前启用新的域名. FFD 是一种典型的短周期恶意域名, 如表 5 所示.

表 5 域名注册信息示例

域名		注册时间	使用年限	过期时间
正常域名	google.com	1997-09-15	14 年	2020-09-14
	www.att.com	1986-04-25	26 年	2018-04-26
	www.paypal.com	1999-07-15	12 年	2018-07-15
FFD 域名	ricalt.com	2012-02-04	3 月	2013-02-05
	datetop.ru	2011-10-12	7 月	2012-10-12
	www.electro-kur.ru	2011-03-12	1 年	2013-03-12

2) TTL 值

TTL 是域名解析记录在本地域名服务器中的存留时间[12]. Perdisci、Passerini 等分别在研究中指出, TTL 是区分 FFD 和正常域名的主要特征之一[5,13]. 研究指出, 为了减少网络解析时间以更快地提供服务, 正常域名设置较大的 TTL 值; FFD 为了快速轮转替换失效主机设置较小的 TTL 值. 但是, 我们在实际网络数据中发现, 正常服务提供商, 特别是大型网络服务提供商, 为了尽快广播更换服务器信息或尽快交付新上线服务往往会选取较小的 TTL 值以减少服务更换导致的流量损失. 我们的检测数据中大量正常域名 TTL 为 5 秒. 另外, 本地域名服务器具有更改 TTL 的权限[14]导致 TTL 数据失真. 因此, TTL 不能作为区别正常域名和 FFD 的主要特征.

4 特征辨识度分析

本文第 3 节详细讨论了 FFSN 的行为特征, 分析表明 FFSN 原有的标识特征已经不能完全作为 FFSN 的

© 中国科学院软件研究所 http://www.jos.org.cn

行为标识. 本节在 FFSN 行为特征分析基础上评估特征辨识度, 进一步揭示 FFSN 的动态发展性.

Holz 在研究中首次提出了 flux-score 方法用以 FFD 检测[4]. 文中选取整个观测期域名的不同 A 记录数 nA、单次查询 NS 记录数 nNS 以及域名 A 记录所属 AS 数 nASN 三个属性构成特征向量. 我们利用文中训练获取的决策函数公式 1 对实验采集的数据进行验证, 实验结果中误报率(false positive)达到了 100%. 这说明 FFSN 特征已经随着时间推移有了显著变化, 原有基于拟合得到的决策函数已经失效.

$$f(x) = 1.32 \cdot n_A + 18.54 \cdot n_{ASN} + 0 \cdot n_{NS} \quad (1)$$

为了进一步确定 FFSN 的标识特征, 我们依据第 3 小节行为特征相关研究对采集获取的数据进行如表 6 统计处理, 并以此确定 FFSN 的 11 个基础特征, 分别以 $Fi(i \in Z+$且 $1 \leqslant i \leqslant 6)$或 $Fij(i,j \in Z+$且 $1 \leqslant i \leqslant 6)$标记. 根据基础特征构建特征向量, 采用 k-means[15]算法评估不同特征向量对域名类别的辨识度.

表 6 DNS 解析记录处理说明

编号	二级编号	描述
F1	*F11*	统计单次解析中 A 记录数
	F12	统计单次解析中关联 IP 所属 ASN 数
	F13	统计单次解析中关联 IP 所属国家数
F2	*F21*	统计域名所有 A 记录关联的不同 IP 总数
	F22	统计域名所有不同 NS 记录数
	F23	统计域名所有 NA 记录关联的不同 IP 总数
F3	*F31*	统计域名所有关联 IP 所属的 ASN 数
	F32	统计域名所有关联 IP 所属的国家数
F4	-	统计域名所有 A 记录和 NA 记录关联 IP 重叠度
F5	-	统计域名关联 IP 增长情况
F6	-	统计域名关联 IP 的有效性

实验中, 我们选取预知类型的 136 个域名作为样本数据, 其中非 FFD 域名 70 个, FFD 域名 66 个. 聚类分析主要根据指定的特征向量通过计算样本数据与聚类中心的距离将样本划分为两类, 即 FFD 和非 FFD. 为了防止数据量级差异对结果造成影响, 我们对数据进行了归一化处理.

表 7 选取实验中 5 组典型的特征向量辨识结果. 由实验数据特征组 1 和特征组 2 较大的误报率和漏报率可知, 传统的标识特征——关联 IP 总量等已经不能作为 FFSN 的关键特征, FFSN 正在发生新的变化. 对比特征组 5 和其他 4 组特征组, 我们发现辨识的准确度不完全取决于特征量的多少. 表中特征组 4 取得了较好的辨识效果, 这也正符合本文第 3 节分析中指出的 FFSN 的 A 记录和 NA 记录 IP 重叠度、关联 IP 增长率、关联 IP 有效性及 IP 地址分散度与正常服务网络存在较大的差异. 实验表明, 在 FFSN 和正常服务网络的博弈中, FFSN 控制者始终无法规避 FFSN 的固有缺陷, 我们能够通过分析选取有效的标识特征抑制 FFSN 的应用和发展.

表 7 特征向量判别结果

序号	特征向量	FFD	非 FFD	误报	漏报
1	*F1*	91	45	25	0
2	*F2*	42	94	0	24
3	*F3*	61	75	0	5
4	*F4 F5 F6*	67	69	1	0
5	*F1-F6*	56	80	0	10

5 结语

Fast-Flux 服务网络相比以往发生了较大的改变, 原有特征已经不能完全作为 FFSN 的有效标识. 本文对 23000 多个域名向分布在六大洲的 300 个 DNS 服务器发起 DNS 解析请求, 根据解析记录全面讨论 FFSN 的行为特征, 分析 FFSN 新的行为变化. 本文在行为特征分析基础上对行为特征进行辨识度评估, 进一步揭示了 FFSN 的行为变化. FFSN 行为特征研究为恶意域名检测、网络资源保护等多个方面的安全研究提供了依据.

参考文献

1 The Honeynet Project. Know your enemy: Fast-Flux Service Networks, July 2007. http://www.honeynet.org/ papers/ff/.

2 Brisco T. RFC 1794: DNS support for load balancing, April 1995. http://www.ietf.org/rfc/rfc1794.txt.

3 Rzewski P, Day M, Gilletti D. RFC 3570: Content internetworking(CDI) scenarios, July2003. http://www.ietf.org/rfc/rfc3570.txt.

4 Holz T, Gorecki C, Freiling F, Rieck K. Measuring and detecting Fast-Flux Service Networks. Proc. of the 15th NDSS, 2008.

5 Passerini E, Paleari R, Martignoni L, Bruschi D. FluXOR: detecting and monitoring Fast-Flux Service Network. Proc. of the 5th DIMVA, 2008.

6 Huang SY, Mao CY, Lee HM. Fast-Flux Service Network detection based on spatial snapshot mechanism for delay-free

(下转第 33 页)

© 中国科学院软件研究所 http://www.jos.org.cn

用户用键盘点击场景中角色，则角色处于选中状态，然后用鼠标点击目标位置，则输入组件(InputCompont)向逻辑组件(RuleComponent)发送的消息首先存储在 GC 中，GC 在下一帧逻辑处理阶段把该消息发送给逻辑组件(RuleComponent)，逻辑组件经过逻辑计算，计算出了角色下一帧的位置，并把消息传给 GC，GC 在下一帧传给角色(Actor)，则对角色下一帧的位置进行渲染，循环知道角色到达目标地点位置，渲染画面如图 7 所示. 该仿真验证程序的为 60 帧/秒，能够很好的满足人的视觉感受，画面比较流畅.

图 7 角色运动图

6 结语

本文运用设计模式设计了一款具备通用性、灵活性以及可扩展性的面向对象三维游戏引擎. 其主要特点是:体系结构简洁清晰，面向对象技术将引擎划分为多个定义清晰而功能相对单一的子系统，并大大降低各个功能模块的耦合度. 通过合理使用设计模式,使得系统易于修改、维护、升级以及复用，部分模式在一定程度上提高了系统的性能. 目前，该引擎被用于开发空战游戏系统. 该引擎还将在执行效率方面作进一步提升，以适应不断发展而提出的新的要求.

参考文献

1 孙咏.基于 OCP 软件应用架构的设计与实现[博士学位论文].北京:中国科学院研究生院,2009.

2 阎宏.JAVA 与模式.北京:电子工业出版社,2002.41–44.

3 Gamma E.设计模式-可复用面向对象软件基础(双语版).北京:机械工业出版社,2009.1–21.

4 Shalloway A, Trott JR. Design Patterns Explained.第 2 版.北京:人民邮电出版社,2010.53–188.

5 张秀山.虚拟现实技术及编程技巧.长沙:国防科技大学出版社,1999.20–90.

6 Berenbrink P, Brinkman A, Scheduler C. SIMLAB-A Simulation Environment for Storage Area Networks. IEEE, 2001, 27(4):227–234.

(上接第 8 页)

detection. Proc. of the 5th ASIACCS, 2010.

7 Hu X, Knysz M, Shin KG. Measurement and analysis of global IP-usage patterns of Fast-Flux botnets. IEEE INFOCOM technical program, 2011.

8 Caglayan A, Toothaker M, Drapaeau D, Burke D, Eaton G. Behavioral analysis of Fast Flux Service Network. Proc. of the 5th CSIIRW, 2009.

9 汪洋.Fast-Flux 服务网络检测方法研究[硕士学位论文].武汉:华中科技大学,2009.

10 Hawkinson J. RFC1930: Guidelines for creation, selection, and registration of an Autonomous System (AS), March 1996. http://www.ietf.org/rfc/rfc1930.txt

11 Knysz M, Hu X, Shin KG. Good guys vs.bot guise:mimicry attacks against Fast-Flux detection system.IEEE INFOCOM technical program, 2011.

12 Mockapetris P. RFC 1035: Domain name-implementation and specification, November 1987. http://www.ietf.org/rfc/rfc1035. txt.

13 Perdisci R, Corona I, Dagon D, Lee WK. Detecting malicious flux service networks through passive analysis of recursive DNS traces. Proc. of ACSAC'09, 2009.

14 Mockapetris P. RFC 1034: Domain name-concepts and facilities, November 1987. http://www.ietf.org/rfc/ rfc1035.txt.

15 Xiao JZ, Li X. A Research of the Partition Clustering Algorithm. International Symposium on Intelligence Information Processing and Trusted Computing 2010.

© 中国科学院软件研究所 http://www.jos.org.cn